Swadesh Swain

swadesh_s@ece.iitr.ac.in | webpage | Swadesh Swain | Swadesh06 | Swadesh Swain | IITR

Education

**GPA conversion to 4.0 scale done according to: https://www.scholaro.com/app/gpa/report?id=636807*

Indian Institute of Technology Roorkee, Roorkee, Uttarakhand GPA: 7.73/10 (3.62/4.0*)
Bachelor of Technology in Electronics and Communication Engineering *Oct 2022 – Sep 2026*

Selected Awards, Honors & Achievements

HoD Appreciation Letter for the Highest Increase in SGPA (+2.7) in a single semester

Nominated as Institute Representative for NVIDIA AI Summit, Mumbai, India, 2024

All India Rank 1423 in JEE Advanced 2022 | All India Rank 1697 in JEE Mains 2022 out of 1M+ candidates

Publications

Riemannian-Guided Diffusion for Scalable Synthetic Signal Data Generation
Under Review at IJCAI 2026 Jan 2026

- A resource-efficient data generation pipeline for synthetic data generation of vibrational accelerometer signals of ball bearings
- Introduces a lightweight diffusion architecture achieving 99.7% discriminator confidence against real data, trained with only 100 samples per class of real data

CroPA++: Exposing Vulnerabilities in Vision Language Models and Enhancing Adversarial Transferability of Cross-Prompt Attacks
NeurIPS 2025, Reliable ML Workshop | Open Review | PDF | Code Aug 2025

- Introduced CroPA++, a three-fold enhancement to the existing Cross Prompt Attack method

Revisiting CroPA: A Reproducibility Study and Enhancements for Cross-Prompt Adversarial Transferability in Vision-Language Models
TMLR 2025 | *MLRC 2025* | ★ Best Paper Award | Open Review | arXiv | Code June 2025

- Reproduced *"An Image is Worth 1000 Lies: Adversarial Transferability Across Prompts on Vision-Language Models"*, analyzing its limitations in an attempt to further enhance Cross Prompt Attacks for VLMs
- Introduced a novel initialization technique based on providing better attributes to initial noise using Diffusion Models, Improved Cross-Model Transferability, and ASR performance

Riemann Sum Optimization for Accurate Integrated Gradients Computation
NeurIPS 2024, Interpretable AI Workshop | arXiv | Code Dec 2024

- Developed **RiemannOpt**, a framework to optimize sample point selection for Riemann Sum approximations in Integrated Gradients methods
- Reduced computational costs by up to 4x while maintaining similar performance to baseline methods

Projects and Research Work

IEEE Signal Processing Cup, ICASSP 2025 | *Team Member* | Project Report Sep 2024 – Feb 2025

- A team research project aimed at finding optimal methods for advanced Deepfake detection in images, submitted to the IEEE Signal Processing Cup, ICASSP 2025
- Uses modified multi-transform cross attention within EfficientNets in an ensemble.
- Advisor: **Vinod Pankajakshan**, Associate Professor, Vision and Image Processing Lab – IIT Roorkee

Deep RL-enabled Fortnite Agent Oct 2024 – May 2024

- A team project under the Data Science Group, to develop an agent to play the highly complex game "Fortnite", using multi-linear ICVF via temporal difference learning and DreamerV3 as world model
- NVIDIA's COSMOS used as the embedding model to embed video streams for training

Reproducing UPoP | Code Dec 2023 – Feb 2024

- Reproduced results for *"UPop: Unified and Progressive Pruning for Compressing Vision-Language Transformers"*, ICML 2023, for DeiT model, under the BYOP Reproducibility Track 2024
- Conducted Extensive ablation study and introduced enhancements for optimizations
- Achieved 90% model compression with DeiT sustaining a mere 5.5% accuracy drop

Work and Research Experience

University of Maryland, College Park | *Research Collaborator* Nov 2025 – Present
- Counterfactual detection of safety-critical feature suppression & misalignment boundary detection in models
- Developing quantification metric for suppressed features in models, reasoning circuits, and their causal impact on jailbreaks and misaligned behavior
- Advisor: **Dr. Sanghamitra Dutta**, Assistant Professor, University of Maryland, College Park

Meta AI [FAIR] | *Research Collaborator* Sept 2025 – Present
- Benchmarking short-long-term motion anticipation mechanisms and world model knowledge in VideoLMs
- Advisor: **Dr. Koustuv Sinha**, Research Scientist, Meta AI (VJEPA 2 Core Team Author)

Virginia Tech | *Research Intern* Sept 2025 – Present
- Using circuit tracing interpretability methods in modern LLMs to perform refining interventions based on identifiable transcoder features to develop a user-intervenable interactive LLM pipeline
- Advisor: **Dr. Nagender Aneja**, Associate Professor, Virginia Tech

AuraML | *Research Intern* June 2025 – Sept 2025
- Implemented text-to-3D scene generation frameworks with Graph Diffusion Models
- Created a text-to-3D warehouse generator and editor accessible by an extension through Isaac-Sim for industrial simulation applications
- Contributed directly to development of product : AuraSim
- Advisory Lead: **Arjun Gupta**, CTO - AuraML

Robotics Research Centre, IIIT Hyderabad | *Undergraduate Researcher* May 2025 – July 2025
- Conducted research to develop generalized skill-based models in reinforcement learning using Diffusion Models and Sparse MoE frameworks for VLA tasks in multi-Task RL pipelines
- Advisor: **KM Krishna**, Professor and Lab Head, Robotics Research Centre – IIIT Hyderabad

Koita Centre for Digital Health, IIT Bombay | *Machine Learning Engineer* April 2024 – Oct 2024
- Implemented pipelines for RAG-assisted content generation, OCR correction, and content-assisted image reconstruction for medical applications as part of the Digital Health Team
- Directly contributed to the BharatGen consortium project
- Project Chair Professor: **Ganesh Ramakrishnan**, Koita Centre for Digital Health – IIT Bombay; Supervisor: **Kundeshwar Pundalik**, Sr. Generative AI Engineer

C.A.N.D.L.E Research Lab, IIT Roorkee | **BOSCH** | *Undergraduate Researcher* Dec 2023 – Present
- Conducted research in collaboration with BOSCH on geometric manipulation of latent manifolds of Diffusion Models for editing, conditioning, and interpretability
- Synthetic Data Generation using Inference-time Pullback-based Latent Space Conditioning of Diffusion Models
- Previous research experience includes Multi-Modal model compression, and image processing
- Advisor: **Dr. Sparsh Mittal**, Associate Professor – IIT Roorkee, Bosch Project Advisor: **Dr. Thakre Sanket Sanjay** - Bosch India

Leadership & Activities

Transactions on Machine Learning Research | **TMLR** | **JMLR** | *Reviewer* Sep 2025 – Present
- Contributing to the reviewing discourse of submitted research works at TMLR

Data Science Group, IIT Roorkee | *Joint Secretary* Jan 2024 – Present
- Mentoring students to publish research at several A* conferences and notable peer-reviewed venues such as NeurIPS, CVPR, ICLR, with most works led solely by undergraduate authors
- Currently leading the student AI group. Contributing to several projects in AI Safety and Interpretability
- Undertook and organized lectures, workshops, hackathons spanning the field of Data Science and AI

International Relations Cell | *Joint Secretary* Sep 2023 – May 2025
- Created database of research programs and scholarships abroad available for IIT Roorkee students
- Organizing various talk/interview events with partnerships from universities and students from abroad

Eco Group | *Joint Secretary* | *Head of Editorial* May 2023 – May 2025
- Led sustainability efforts, including a campus-wide plastic cutlery ban, e-waste drives and Motor-Vehicle Free Days

Unnat Bharat Abhiyan | *Executive Member* Sept 2023 - Sept 2024
- Led Science Lab Setup initiatives through UBA, establishing hands-on learning facilities across 5 rural villages